ECGGAN: A Framework for Effective and Interpretable Electrocardiogram Anomaly Detection

Huazhang Wang
Beijing Institute of Technology
P. R. China
huazhang_wang@bit.edu.cn

Zhaojing Luo
National University of Singapore
Singapore
zhaojing@comp.nus.edu.sg

James W. L. Yip
National University Heart Centre
Singapore
james_yip@nuhs.edu.sg

Chuyang Ye
Beijing Institute of Technology
P. R. China
chuyang.ye@bit.edu.cn

Meihui Zhang*
Beijing Institute of Technology
P. R. China
meihui_zhang@bit.edu.cn

ABSTRACT

Heart is the most important organ of the human body, and Electrocardiogram (ECG) is an essential tool for clinical monitoring of heart health and detecting cardiovascular diseases. Automatic detection of ECG anomalies is of great significance and clinical value in healthcare. However, performing automatic anomaly detection for the ECG data is challenging because we not only need to accurately detect the anomalies but also need to provide clinically meaningful interpretation of the results. Existing works on automatic ECG anomaly detection either rely on hand-crafted designs of feature extraction algorithms which are typically too simple to deliver good performance, or deep learning for automatically extracting features, which is not interpretable.

In this paper, we propose ECGGAN, a novel reconstruction-based ECG anomaly detection framework. The key idea of ECGGAN is to make full use of the characteristics of ECG with the periodic metadata, namely beat, to learn the universal pattern in ECG from representative normal data. We establish a reconstruction model, taking leads as constraints to capture the unique characteristics of different leads in ECG data, and achieve accurate anomaly detection at ECG-level by combining multiple leads. Experimental results on two real-world datasets and their mixed-set confirm that our method achieves superior performance than baselines in terms of precision, recall, F1-score, and AUC. In addition, ECGGAN can provide clinically meaningful interpretation of results by revealing the extent to which abnormal sites deviate from the normal pattern.

CCS CONCEPTS

• **Applied computing** → **Health care information systems**; • **Human-centered computing** → *Scientific visualization.*

KEYWORDS

ECG Data Analytics; Time Series; Reconstruction-based; Neural Networks; Interpretability

ACM Reference Format:
Huazhang Wang, Zhaojing Luo, James W. L. Yip, Chuyang Ye, and Meihui Zhang. 2023. ECGGAN: A Framework for Effective and Interpretable Electrocardiogram Anomaly Detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), August 6–10, 2023, Long Beach, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3580305.3599812

1 INTRODUCTION

National University Heart Centre (NUHC) is set up by the National University Health System (NUHS) [1] in Singapore to meet the needs of providing better care to the increasing number of patients who are diagnosed with cardiovascular diseases. Cardiovascular diseases seriously threaten human life and health, and they have become the world's primary cause of death with high disability and mortality rate [5, 20, 50]. In NUHC, clinicians work with collaborators from universities and research institutes [38, 43, 47] to develop innovative approaches for cardiovascular disease analytics.

In particular, Electrocardiogram (ECG) is the principal diagnostic tool employed to record the electrical activity of the heart, which is used to diagnose cardiovascular diseases, such as arrhythmias, heart attacks, and heart failure [18, 28, 36]. A standard 12-lead ECG monitors the beating of the heart from 12 angles by electrodes placed in the chest and limbs, including six limb leads (I, II, III, aVF, aVL, and aVR) and six chest leads (V1-V6) [4]. To assist in cardiovascular disease diagnosis, the raw ECG collected from the hospital needs to go through several steps of processing. Figure 1 illustrates the pipeline for automatic anomaly detection for ECG data in NUHC. The general procedure first extracts the ECG of the specified class, such as normal, abnormal (ST-segment elevated (STE), atrial fibrillation (AF), premature atrial contraction (PAC), etc.), from the database. Then the physiological and environmental noises, e.g., baseline wander, power line interference, and muscle artifacts, are filtered from the ECG to ensure data quality. Subsequently, algorithms are designed based on medical knowledge to extract key features or patterns in the ECG, such as P wave, RR interval, ST slope, and determine whether they exceed the normal value range

[1] NUHS is an integrated academic health system that operates four public hospitals and a number of national specialty centers and polyclinics in Singapore.

*Corresponding author

KDD '23, August 6–10, 2023, Long Beach, CA, USA

ACM ISBN 979-8-4007-0103-0/23/08…$15.00
https://doi.org/10.1145/3580305.3599812

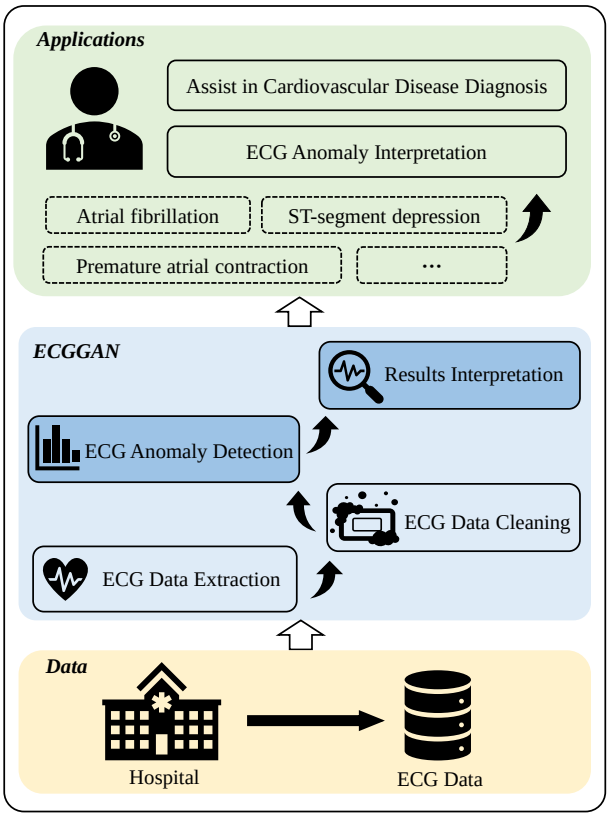


Figure 1: The overview of ECGGAN.

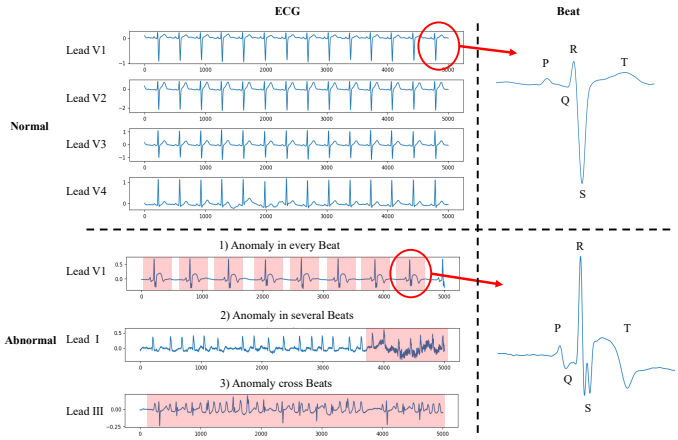


Figure 2: Normal and abnormal ECG examples. Anomalies are highlighted with red background color. Note that full 12-lead ECG include lead I, II, III, aVR, aVL, aVF, and V1-V6. Due to space constraint, we only show few leads as examples.

to detect ECG anomalies. Finally, we can provide interpretable results to clinicians by locating the specific leads and sites of the anomalies [3, 54, 58]. Performing automatic anomaly detection for the ECG data is challenging because we not only need to accurately detect the anomalies but also need to provide clinically meaningful interpretation of the results [26, 59, 61]. To address the above two challenges, we propose ECGGAN, a reconstruction-based ECG anomaly detection framework for effective and interpretable cardiovascular disease analytics, which focuses on the ECG Anomaly Detection and Results Interpretation steps in Figure 1.

To be specific, the heart has a highly consistent physiological structure, and the ECG signals generated by the beating of the heart are also consistent [32]. Normal heart activities are therefore associated with a common normal ECG pattern. As shown in Figure 2, a normal ECG consists of multiple normal periodic metadata, known as beat, which has similar waveform structures, including P, QRS, and T waves [23, 42] (shown in the upper right of the figure). Abnormal ECG signals can be grouped into three broad categories according to the clinical features. First, the anomalies can be observed in every beat in the lead, such as STE. An example can be seen in the first abnormal ECG at the bottom left of Figure 2, where we highlight the abnormal beats with red color. Second, the anomalies are present only within several beats. An example is AF shown in example 2) at the bottom left of Figure 2, where the anterior beats are normal but each of the posterior beats is abnormal. Third, the anomalies are reflected across multiple beats, which can only be detected together with the anterior and posterior beats. The example 3) at the bottom left of in Figure 2 shows the ECG of PAC, where the intervals of beats are irregular. Moreover, the ECG characteristics of different leads present significant differences, which can be observed at the top left of Figure 2 where we only show four (out of 12) leads (i.e., V1-V4 for normal ECG), posting another challenge for ECG anomaly detection and localization.

Most existing methods for ECG anomaly detection extract features from the original ECG, and subsequently determine if they are abnormal by directly comparing with the threshold set in medical machines or feed them into various machine learning models such as SVM [49, 73] and KNN [57, 70] for classification. For example, Philips DXL [76] extracts heart rate, P wave and other features to detect anomalies by determining whether they exceed the threshold, e.g., heart rate over 100 beats per minute or P wave width over 2.5mm. However, these conventional methods are simplistic and rely heavily on hand-crafted feature extraction, which cannot adapt well to complex ECG anomalies. With the application of deep learning (DL) in healthcare [12, 27, 44–46, 48, 69, 71], deep neural networks (DNNs), especially convolutional neural networks (CNNs) [53, 56, 68, 74], have been widely used in ECG anomaly detection for better predictive performance. However, the performance of DNNs highly depends on the quantity and quality of annotated data, which is typically difficult to obtain in large quantity due to data privacy protection regulations [9, 55]. Additionally, clinicians often find it challenging to trust complex black box models due to their poor interpretability, which further raises the barrier to the adoption of DNNs in healthcare.

Since multi-lead ECG can be regarded as multivariate time series, some other works take advantage of reconstruction-based time series anomaly detection methods in general domain [2, 11, 17, 21, 35, 62] for ECG data. These reconstruction-based methods typically generate output data from the input data based on the general pattern learned from the training data. Then, they calculate the difference between the original data and the generated data as the anomaly score and compare with the hand-crafted threshold to

identify the anomalies. In general domain, however, anomalies in the time series may just be "fluctuations", which are points deviated from most of the data. In contrast, ECG data are composed of periodic fluctuations (i.e., beat), and the anomalies can be (1) points that deviate from normal fluctuations or (2) major abnormal patterns with minor normal patterns. As a result, these methods are not suitable for ECG anomaly detection as they may misinterpret normal fluctuations in ECG as anomalies or show insensitivity when identifying minor anomalies. For ECG, Zhou et al. [75] propose BeatGAN, a reconstruction-based model using generative adversarial network (GAN) [22]. This method performs anomaly detection in units of beats rather than the entire ECG, and thus can only detect certain beat-level anomalies. Recall the three types of anomalies in Figure 2, BeatGAN performs well in detecting the first type anomalies, but may fail for the other two types (i.e., anomalies like AF that appear in only several beats or anomalies like PAC that are present across beats) since it processes every single beat independently and never relates them even for the beats in the same lead. Moreover, it also ignores the information from multiple leads, which is important for identifying ECG anomalies.

In this paper, we propose ECGGAN for effective and interpretable cardiovascular diseases analytics. The key idea of ECGGAN is to learn the universal pattern from representative normal data based on periodic and regular fluctuations, and then reconstruct ECG to measure the difference between anomalies and normal patterns. Unlike BeatGAN, our method is able to detect anomalies beyond the beat-level and exploits the information from multiple leads. To be specific, since it is difficult for the reconstruction model to capture every detail in the ECG and properly reconstruct the long ECG signal, we design an ECG cutting module that cuts ECG into periodic metadata (as beats) to learn the detailed pattern in ECG in a finer-grained manner. In reconstruction module, the reconstruction of beat-level signals is achieved with a conditional GAN (cGAN), where the input beat-level signal is conditioned on the lead index as the constant. Subsequently, we develop a mechanism, namely ECG restoring module, which restores the reconstructed beats with lead constraint to obtain the complete reconstructed ECG signals. Finally, to detect the anomalies in ECG, the residual matrix between the reconstructed ECG and the original ECG is fed into the anomaly detection module, which comprises multi-scale filters and the Convolutional Block Attention Module (CBAM) [67] to capture the relationship between different leads. In this way, ECGGAN enables to assist in the diagnosis of cardiovascular disease. Besides accurate anomaly detection, our method provides interpretability of results in the anomaly interpretation module, by visualizing the residual matrix through heat map, which locates the specific sites and leads of the anomalies. We summarize the main contributions as follows.

- We propose ECGGAN, a novel reconstruction-based ECG anomaly detection framework to automatically detect anomalies and provide interpretable results for ECG data.
- We study the characteristics of ECG, and design a cutting, reconstruction and restoring mechanism for ECG data to learn the fine-grained normal pattern of beats and reconstruct the complete ECG.
- We propose a conditional GAN, which takes leads as constraints to capture the unique characteristics of each lead, and construct an anomaly detection module with multi-scale filters and CBAM to detect anomalies at ECG-level.
- We conduct extensive experiments on two real-world ECG datasets and their mixed-set. The results confirm that our method achieves superior detection performance in terms of precision, recall, and F1-score.

In the remainder of this paper, we review the existing works in Section 2. In Section 3, we present the problem formulation, and introduce the design and implementation details of each module of ECGGAN. Section 4 shows the experimental results and the interpretability of ECGGAN. Finally, we conclude the paper and describe future directions in Section 5.

2 RELATED WORK

Conventional Feature Extraction-based Anomaly Detection. Programs for computer-aided automated ECG analysis have been available since the 1950s. Early methods generally extract ECG features (P wave, QRS wave, T wave, etc.) based on hand-crafted designs, such as DWT [14, 30, 49, 57], CWT [30], DCT [30], and Db4 [15, 52], and then detect anomalies by deciding whether the value or the shape deviates from the normal range [8, 16, 34, 60]. For example, Philips DXL [76] analyzes single-lead or 12-lead ECG automatically by identifying pacemaker pulses, P wave, QRS wave; the Marquette 12SL algorithm [19] uses chest leads to extract heart rate, electrical axis and other features for arrhythmia detection; HES (Hannover ECG System) [51] can detect anomalies in limb and chest according to hand-crafted setting thresholds. In more advanced works, the extracted features are fed into machine learning classifiers, such as SVM [49, 73], Decision Tree or Random Forest [37, 72], and KNN [57, 70], for ECG anomaly detection.

The above methods rely on expert knowledge to design the algorithm for feature extraction, rather than fully automatic feature extraction. Moreover, these conventional methods can be simplistic and may not adapt well to the complex clinical environment, such as excessive noise, multi-lead characteristics, and variability of diseases and patients.

Deep Learning-based Anomaly Detection. In recent years, DL-based methods have been applied to ECG anomaly detection. Since multi-lead ECG data is similar to images, CNNs are often used for the task [7, 53, 56, 66, 68, 74]. AY et al. [24] propose a DNN that classifies 12 cardiac rhythm categories, and the average F1-score is higher than general medical experts, which verifies the effectiveness of the end-to-end DL-based method for heart rhythm classification. Liu et al. [41] propose a ResNet architecture [25] with an attention mechanism, which assigns weights to each local feature of ECG by its importance, and automatically optimizes the weights during model training. Kwon et al. [33] propose an Ensemble Neural Network, which combines demographic information, extracted ECG features, and feature vectors from the original ECG to diagnose the left ventricular hypertrophy disease. However, these DL-based methods are black-box models, which have poor interpretability and cannot provide reliable reference for doctors' clinical diagnosis.

In order to provide interpretable results, many anomaly detection works take advantage of reconstruction-based methods. Autoencoder (AE) [29] is one of the most classical reconstruction models,

which is composed of Encoder and Decoder. However, AE is data-dependent, which can only process data similar to the training data and cannot extract features beyond the training data well. To address the limitations of AE, Variational Autoencoder (VAE) [10] learns the distribution of training data so that it can generate more comprehensive data. In addition, reconstruction methods based on GAN [22] are also the current main research direction. Akcay et al. [2] propose GANomaly, with the generator consists of Encoder1-Decoder-EnCoder2, and use the difference of hidden space features produced by Encoder1 and Encoder2 as the reference for detecting anomalies. Considering the characteristics of time series, Li et al. [35] take Long Short-term Memory (LSTM) as the basic network of GAN, namely MAD-GAN, using a novel Discrimination and Reconstruction Anomaly Score (DR-Score) to combine the two losses to detect potential anomalies. However, the above general reconstruction-based methods only focus on time series with normal data mostly, and lack adjustments specific of ECG data, which leads to an unsatisfying performance. BeatGAN [75], a GAN-based method proposed by Zhou et al., is able to detect abnormal rhythms at the beat-level by reconstructing the data and calculating the error between the original data and the reconstructed data as an anomaly score. However, BeatGAN only focuses on beat data, and cannot exploit the information cross beats and the unique characteristics of lead for detecting anomalies at ECG-level.

3 METHODOLOGY

In this section, we first formulate the problem of ECG anomaly detection and interpretation. Subsequently, we present the overview of ECGGAN and elaborate on each of its modules. We further discuss how to give an interpretable result based on the reconstructed ECG. Table 1 lists the definitions of symbols used in this paper.

3.1 Problem Formulation

Multi-lead ECG represents electrical signals recorded from different angles, reflecting the health of different regions of the heart, e.g., lead I is composed of the signals recorded by the sensors placed on wrists. Figure 2 shows a portion of the standard 12-lead ECG. The diagnosis of cardiovascular disease is often made by one or more specific leads, e.g., ST-segment depression (STD) caused by subendocardial ischemia is usually present in leads V4-6, I, II, and aVL. The task studied in this paper is to take multi-lead ECGs as input and detect the anomalies present in the ECG signals.

Formally, let $X \in \mathbb{R}^{M\times N}$ be a multi-lead ECG signal, which has M leads for each time tick t and N time ticks in length in one lead. beat is the periodic metadata of ECG in a lead. The j-th beat in the i-th lead is denoted by $x_{ij} \in \mathbb{R}^{1\times L}$, where L is the beat length. Accordingly, we have $X = \{x_{ij} | i = 1, \cdots, M; j = 1, \cdots, k\}$, where k is the number of beats in one lead of ECG. Anomaly is defined as the segment that deviates significantly from the normal pattern. Given a collection of multi-lead ECGs, $\mathcal{X} = \{X_1, X_2, \cdots\}$, our *anomaly detection task* is to output the binary labels, $Y \in \{0, 1\}$, to indicate whether the corresponding ECG is normal or abnormal. *Anomaly interpretation* is to explain why the ECG is identified as abnormal, i.e. to provide the specific sites, leads and severity of the anomalies in the ECG.

Table 1: Table of symbols

Symbol	Definition
M	number of leads in ECG
N	length of ECG
$X \in \mathbb{R}^{M\times N}$	multi-lead ECG
Y	label of ECG
L	length of beat
$x_{ij} \in \mathbb{R}^{1\times L}$	the j-th beat in the i-th lead of ECG
y	lead index of beat
k	number of beats in one lead
$\mathcal{X} = \{X_1, X_2, \cdots\}$	collection of multi-lead ECGs
$Rlist$	list of beats' positions
t	the time tick index of ECG
L_{avg}	average length of beats
X'	reconstructed ECG
x'	reconstructed beat
$\mathbf{M}$	residual matrix
$G(\cdot)$	generator
$D(\cdot)$	discriminator
$G_E(\cdot)$	encoder of $G(\cdot)$
$G_D(\cdot)$	decoder of $G(\cdot)$
$D_E(\cdot)$	encoder of $D(\cdot)$
L_G	loss of $G(\cdot)$
L_D	loss of $D(\cdot)$
$L_{G-label}$	label loss item of $G(\cdot)$
L_{G-gen}	generation loss item of $G(\cdot)$
$L_{G-feature}$	feature loss item of $G(\cdot)$
λ_{label}	weight of $L_{G-label}$
λ_{gen}	weight of L_{G-gen}
$\lambda_{feature}$	weight of $L_{G-feature}$
$baselineDrift$	the baseline drift error in ECG
$\mathbf{d}$	kernels sizes of multi-scale filters
F	latent features
$\sigma(\cdot)$	sigmoid function
lr	learning rate
r	dropout radio

3.2 Framework

The framework of ECGGAN is given in Figure 3. ECGGAN comprises four major modules: 1) The *ECG Cutting module* cuts the original ECG into periodic metadata beats, retaining lead indexes and cutting ticks; 2) The *Reconstruction Model with Lead Constraint* learns the normal pattern with the unique characteristics of each lead and reconstructs beats; 3) The *ECG Restoring module* restores the complete reconstructed ECG signal from the reconstructed beats; 4) The *Anomaly Detection and Interpretation module* detects the deviations from normal pattern and locates these anomalies.

In the training process, we first cut the normal ECGs into beats with the ECG cutting module, and then train the reconstruction model with leads as constraints to learn the universal normal pattern (see the left part of Figure 3). At test time, given an ECG X, we input the beats obtained from the cutting module into the trained model to obtain the reconstructed beats and the reconstructed ECG X' by the lead indexes and cutting ticks. Finally, the residual matrix $\mathbf{M}$ between X and X' is fed into the anomaly detection module to

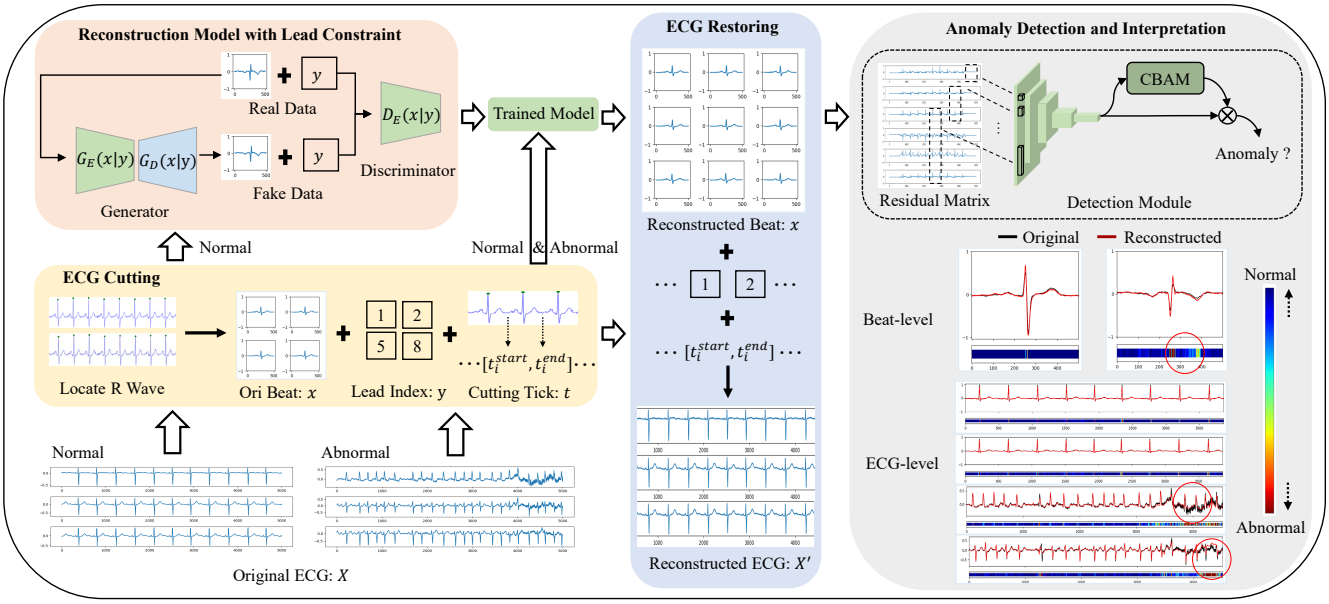


Figure 3: The framework of ECGGAN

detect anomalies, as shown at the upper right of Figure 3. Besides anomaly detection, we visualize **M** to locate and indicate the severity of the anomalies by revealing the abnormal deviations from the normal pattern, as shown at the bottom right of Figure 3.

3.3 ECG Cutting Module

Since it is challenging to directly reconstruct the complete ECG signals, the ECG cutting module cuts ECG signals into beats, which can be uniquely determined by the largest peak, namely the R wave. First, we detect the R wave ticks through heartpy [64, 65] based on lead I to locate the position of beats, denoted by $Rlist$, where $Rlist[i]$ represents the time tick of the R wave of the i-th beat in the whole ECG. Then, all signals in the other leads are cut into beats with the same ticks, as these signals are aligned in the time domain.

To accurately cut each ECG and adapt the various length of beats from different patients, we calculate the average length L_{avg} of the distance between adjacent R wave ticks and use this length for ECG cutting. For the i-th beat, we can pinpoint the start index t_i^{start} and end index t_i^{end} as

$$t_i^{start} = Rlist[i] - \frac{L_{avg}}{2}, t_i^{end} = Rlist[i] + \frac{L_{avg}}{2}. \quad (1)$$

For every beat, we also retain the lead index, denoted by y, as well as the start and end time tick t^{start}, t^{end}. All beats are zero-padded or cropped to a uniform length L. In addition, we remove the beats at the left and right ends of the raw ECG as they usually do not contain a complete period (see the middle left of Figure 3).

3.4 Reconstruction Model with Lead Constraint

After cutting the ECG signals into beats, we construct a model that takes a beat as input and seeks to reconstruct the original beat for normal data, where the lead index is also included as a constraint. This allows the model to capture the unique characteristics in each lead and learn the normal pattern in the beat. The model is trained with normal data, and when abnormal data is fed into the trained model, the output is expected to differ from the input signal, especially for the parts with abnormality. Figure 4 shows the architecture of the reconstruction model with the lead constraint. The reconstruction model comprises a generator and a discriminator, and their detailed designs are described below.

Generator $G(\cdot)$ learns the universal pattern from large-scale normal beats. The input of $G(\cdot)$ is a beat x, and its output is denoted by x' with normal pattern, as shown in the upper part of Figure 4. $G(\cdot)$ uses an autoencoder, which is composed of an encoder $G_E(\cdot)$ and a decoder $G_D(\cdot)$. During training, the goal of $G(\cdot)$ is to generate beats with normal pattern and deceive $D(\cdot)$ to identify the generated samples as real. The generator needs to produce patient-specific information and the unique lead characteristics in x. So, it takes the ECG signal x and lead information y into account, and its optimization objective is to minimize the following loss $L_{G-label}$:

$$L_{G-label} = E_{x\sim P_x}\left[\log(1 - D(G(x|y)))\right], \quad (2)$$

where $D(G(x|y))$ represents the decision of the discriminator for the output of the generator based on x and y. Specifically, we feed beat and the one-hot encoding of y into the 1-D convolution layer with similar structure, and concatenate the output feature vectors to obtain the joint representation, shown as the left part in Figure 4.

However, it is not sufficient to rely only on $L_{G-label}$ to learn the normal pattern. Even though x' has a universal normal pattern, x' may be morphologically dissimilar from x, leading to incorrect identification of unimportant deviations as anomalies. Therefore, to establish the explicit association between x and x', we add L_{G-gen} to $G(\cdot)$ as a regularization term to measure the morphological differences between x and x':

$$L_{G-gen} = E_{x\sim P_x}\left[\left\|x - x'\right\|_2^2\right]. \quad (3)$$

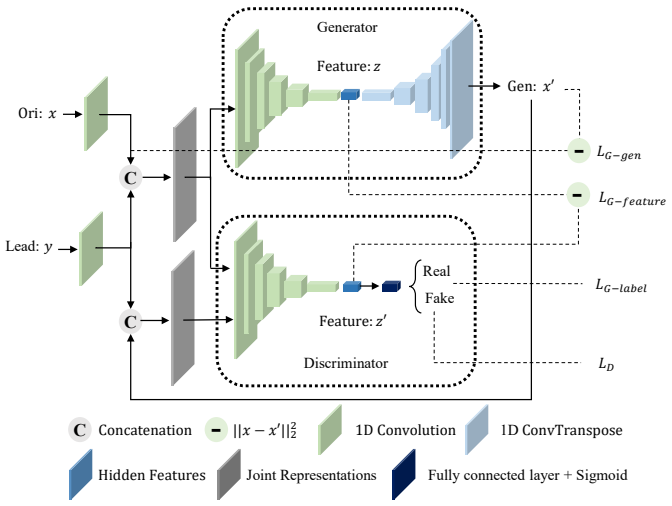


Figure 4: The architecture of the reconstruction model

In addition, the differences between x and x' in the latent space also contain the data distribution characteristics, which is beneficial for a better reconstruction. Therefore, we design $L_{G-feature}$ as another regularization term to measure the latent feature differences and establish the implicit associations between x and x':

$$L_{G-feature} = E_{x\sim P_x}\left[\left\|f_D(x|y) - f_G(x'|y)\right\|_2^2\right], \tag{4}$$

where $f_G(\cdot)$ and $f_D(\cdot)$ represent the latent feature vectors from $G(\cdot)$ and $D(\cdot)$, respectively.

The overall optimization objective of $G(\cdot)$ consists of three terms, $L_{G-label}$, L_{G-gen} and $L_{G-feature}$:

$$\begin{aligned} L_G =& \lambda_{label} * L_{G-label} + \lambda_{gen} * L_{G-gen} \\ &+ \lambda_{feature} * L_{G-feature}, \end{aligned} \tag{5}$$

where λ_{label}, λ_{gen} and $\lambda_{feature}$ are hyper-parameters to represent the weights of the loss terms.

Discriminator $D(\cdot)$ is composed of an encoder $D_E(\cdot)$ with the same structure of $G_E(\cdot)$, followed by a fully connection layer and a sigmoid function, as shown in the bottom part of Figure 4. Given an beat with lead index, $D(\cdot)$ determines whether the beat is real or fake, and its optimization objective is to maximize the loss L_D:

$$L_D = \mathbb{E}_{x\sim p_x}\left[\log D(x|y)\right] + \mathbb{E}_{x\sim p_x}\left[\log(1 - D(G(x|y)))\right]. \tag{6}$$

where the first and second terms encourage $D(\cdot)$ to recognize real data and identify fake data, respectively. $D(\cdot)$ and $G(\cdot)$ are trained simultaneously with a two-player min-max game, and the trained $G(\cdot)$ learns the normal pattern in the training data.

3.5 ECG Restoring Module

The ECG restoring module aims to restore the complete reconstructed ECG signal from the reconstructed beats based on lead indices and cutting ticks, so that information at different locations can be jointly exploited for ECG-level anomaly detection.

First, we zero-pad or crop the reconstructed beats of the uniform length L to the original size based on the R wave and cutting ticks:

$$t_i^{start-beat} = \frac{L}{2} - \left(Rlist[i] + t_i^{start}\right), \tag{7}$$

$$t_i^{end-beat} = \frac{L}{2} + \left(t_i^{end} - Rlist[i]\right). \tag{8}$$

where $t_i^{start-beat}$ is the start time tick in the i-th reconstructed beat, and $t_i^{end-beat}$ is the end time tick in this beat. Then, we restore the reconstructed ECG X' by stitching the reconstructed beats of the original size according to the lead indices and cutting ticks, which enables X' to retain the original location and lead features.

In addition, there is usually baseline drift caused by power line interference and muscle noise in ECG, which makes ECG signal offset [63]. To eliminate the baseline drift, we calculate the average value at both ends of all beats of each lead:

$$baselineDrift = \frac{1}{2k}\sum_{i=1}^{k}\left(x_i[t_i^{start}] + x_i[t_i^{end}]\right). \tag{9}$$

Then, we subtract the baseline drift from the original and reconstructed ECG.

3.6 Anomaly Detection and Interpretation

In view of the clinical requirement of accuracy and interpretation, we design anomaly detection and interpretation modules to detect and locate the anomalies by measuring the deviation of abnormal sites from normal pattern, as shown in the right part of Figure 3.

Anomaly Detection. A straightforward idea like the general reconstruction-based method is using the errors between the original ECG X and the reconstruction X' as the score for anomaly detection directly, but it is hard to detect minor anomalies and the location information is ignored. Consequently, to measure the error between ECG and normal pattern accurately, we first define the residual matrix $\mathbf{M}$ in anomaly detection module:

$$\mathbf{M} = \left|X - X'\right|. \tag{10}$$

Since $\mathbf{M}$ can be viewed as an image, we design the anomaly detection module by classifying $\mathbf{M}$ with a 2D CNN, as shown in the upper right of Figure 3. As $\mathbf{M}$ corresponds to multiple leads, the 2D CNN comprises multi-scale filters [13], so that the joint information of adjacent leads is captured. Specifically, we preset a set of multi-scale filters, and the kernels sizes of them are denoted by $\mathbf{d}$, where $d[i]$ represents the kernel size of the i-th filter in the lead dimension. Each filter is associated with an inception block and takes $\mathbf{M}$ as input. The outputs of these filters are concatenated passed through sequential 2D convolutional layers, which gives the feature maps F that are to be fed into the classification layers.

To better detect anomalies, the classification layers combine the channel attention mechanism and spatial attention mechanism, which is achieved with CBAM [67]. Specifically, for features F of size $[C, H, W]$, where C is the channel number, H is the height, and W is the width of the features, channel attention first conducts average pooling and max pooling operations for each channel. The results are then processed by a shared Multi-Layer Perception (MLP) and merged. Finally, the sigmoid function is used for the final output of channel attention. Mathematically, the computation $M_c(F)$ of

channel attention described above is represented as

$$M_c(F) = \sigma\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right), \quad (11)$$

where $\sigma(\cdot)$ is sigmoid function and the dimension of $M_c(F)$ is $[C, 1, 1]$.

Then, F is multiplied by M_c to obtain F_c for computing spatial attention. Similarly to channel attention, spatial attention first compresses the channels by conducting average pooling and max pooling along the channels. The two pooling results are concatenated and fed into a 7×7 convolutional layer followed by a sigmoid function, which produces the spatial attention. Mathematically, the computation $M_s(F)$ of spatial attention is represented as

$$M_s(F_c) = \sigma\left(f^{7\times7}\left(\left[AvgPool\left(F_c\right); MaxPool\left(F_c\right)\right]\right)\right), \quad (12)$$

where $f^{7\times7}(\cdot)$ represents the 7×7 convolution layer, $[\cdot;\cdot]$ is concatenation operation, and the dimension of $M_s(F_c)$ is $[1, H, W]$.

Finally, we multiply F_c and M_s to utilize the spatial attention, and pass the result through a fully connected layer and a softmax activation function to output the abnormality likelihood of X.

Anomaly Interpretation. In clinical diagnosis, doctors usually pay special attention to the sites that deviate from normal pattern, and diagnose the lesions in heart according to the anomalies. M_{ij} (the (i, j)-th entry of $\mathbf{M}$)) represents the deviation of the reconstructed ECG X' from the original ECG X at lead i and time j, and a greater M_{ij} represents more severe anomaly. Therefore, we visualize M and highlight the abnormal sites by a heat map, as shown in the bottom right part of Figure 3. The color is close to blue when M_{ij} is small, whereas the color close to red means that X deviates from the normal pattern seriously at the corresponding lead and location. For normal ECG, the heat map tends to be blue, indicating that it is less likely to have abnormality. For abnormal ECG, there will be obvious red or yellow areas in the visualization, which means that these sites deviate from the normal pattern and are very likely to have anomalies.

4 EXPERIMENTS

In this section, we evaluate the effectiveness of ECGGAN, and demonstrate the interpretability in locating ECG anomalies.

4.1 Experimental Setup

4.1.1 Datasets. While we are developing our ECGGAN framework, the data engineers in NUHC are still in the process of cleaning and extracting the raw ECG data from NUHS hospitals. Consequently, suggested by clinicians from NUHC, we evaluate the effectiveness of ECGGAN using publicly available real-world ECG datasets from other hospitals first, before we deploy ECGGAN in NUHC. The details of the two real-world ECG datasets are as follows:

- **CPSC [39]** contains ECGs collected from 11 hospitals, including normal data and 8 categories of diseases such as AF and PAC. Each sample is a standard 12-lead ECG signals with a sampling frequency of 500Hz. We select 932 high quality 12-lead ECGs with 466 normal and 466 abnormal ECGs.
- **AIWIN [1]** contains normal and abnormal ECGs, which are composed of standard 12-lead ECG signals with a sampling frequency of 500Hz. Similarly, due to data quality problems, we get a total of 768 ECGs after cleaning the dataset.
- **Mixed-Set** contains 1700 ECGs with 850 normal data and 850 abnormal data. Since the data in CPSC and AIWIN are 500Hz standard 12-lead ECG, we mix them as an additional verification set for our proposed method.

4.1.2 Evaluation Metrics. We evaluate the effectiveness of ECGGAN with the metrics of precision, recall, F1-score and Area Under Curve (AUC) with the standard deviation. We split the normal and abnormal ECGs in the dataset into 10 folds with equal probability and perform 8: 1: 1 split for training, validation, and testing, respectively. We take each fold of data as a test independently to cover all ECGs, calculate ten times and take the average as the test result.

4.1.3 Baseline Methods. We compare ECGGAN with three categories of ECG anomaly detection baselines: (1) feature extraction-based methods: SVM [49] and KNN [57]; (2) general classification-based methods: CNN [24]; (3) general reconstruction-based methods: LSTM [40], VAE [31], GAN [22], MAD-GAN [35], GANomaly [2], and BeatGAN [75]. We briefly introduce these baselines as follows.

- SVM [49] uses the extracted features based on PCA method for classification and automatic diagnosis of anomalies.
- KNN [57] finds the k-nearest neighbors and classifies the ECG by neighbor category.
- CNN [24] is the general classification-based model that extracts the features of ECG for anomaly detection.
- LSTM [40] takes LSTMs as the backbone network and establishes an autoencoder model to detect the anomaly.
- VAE [31] is a reconstruction model that uses the Variational Autoencoder to learn ECG latent variable information.
- GAN [22] is a general reconstruction model consisting of the generator and the discriminator.
- MAD-GAN [35] considers the entire variable set concurrently to capture the latent interactions amongst the variables for Multivariate Anomaly Detection.
- GANomaly [2] uses a conditional generative adversarial network that jointly learns the generation of high-dimensional space and the inference of latent space.
- BeatGAN [75] is an anomaly detection algorithm for anomalous beats based on adversarially generated time series.

4.1.4 Experimental Settings. We implement ECGGAN with Pytorch library and set beat length L to 500. The encoder $G_E(\cdot)$ in beat-level Reconstruction model is composed of five 1D convolution layers, followed by BatchNorm and LeakyRelu function with the leak slope to 0.2. The kernels of each layer in $G_E(\cdot)$ are 128(4/2/5)-256(4/2/1)-512(4/2/1)-1024(4/2/1)-1024(4/2/1), where 128(4/2/5) means the number of filters is 128, the size of filter is 4, the stride is 2 and the padding is 5. The decoder $G_D(\cdot)$ is composed of six 1D convolution layers similar to $G_E(\cdot)$. We optimize the training process similar to WGAN [6], setting the clamp value of $D(\cdot)$ parameter update to be 0.01, and the critic rounds of $G(\cdot)$ to be 2. Moreover, the learning rate $lr = 0.005$, dropout rate $r = 0.5$, $\lambda_{label} = 1, \lambda_{gen} = 1, \lambda_{feature} = 1$. Anomaly Detection module is composed of a multi-scale 2D convolution layer in inception, followed by four 2D convolution layers and the residual matrix $\mathbf{M} \in \mathbb{R}^{12\times5000}$. We design seven filters of different sizes, and the kernel's size is $[(\mathbf{d}, 40)/(1, 4)]$, while $\mathbf{d}$ is [1, 2, 3, 4, 6, 8, 12].

Table 2: Performance of ECGGAN and baselines

Method	CPSC				AIWIN				Mixed-Set			
	Precision	Recall	F1-score	AUC	Precision	Recall	F1-score	AUC	Precision	Recall	F1-score	AUC
SVM	0.519±0.077	0.987±0.120	0.680±0.006	0.645±0.042	0.625±0.053	0.649±0.085	0.637±0.012	0.581±0.060	0.517±0.050	0.947±0.107	0.669±0.006	0.603±0.034
KNN	0.501±0.005	**1.000±0.000**	0.667±0.004	0.503±0.007	0.541±0.028	0.805±0.063	0.647±0.006	0.559±0.027	0.503±0.016	0.912±0.029	0.648±0.006	0.524±0.033
CNN	0.748±0.057	0.936±0.023	0.832±0.032	0.807±0.042	0.769±0.057	0.893±0.042	0.826±0.025	0.807±0.035	0.766±0.031	0.911±0.034	0.832±0.018	0.814±0.023
LSTM	0.625±0.019	0.933±0.016	0.749±0.017	0.750±0.037	0.539±0.040	0.919±0.070	0.680±0.011	0.558±0.033	0.565±0.032	0.931±0.036	0.704±0.018	0.639±0.042
VAE	0.594±0.024	0.951±0.045	0.731±0.015	0.673±0.042	0.512±0.011	0.977±0.031	0.672±0.006	0.532±0.055	0.545±0.029	0.948±0.053	0.693±0.012	0.606±0.047
GAN	0.615±0.021	0.930±0.027	0.743±0.014	0.680±0.039	0.515±0.011	0.980±0.018	0.676±0.006	0.545±0.041	0.543±0.032	0.952±0.051	0.691±0.014	0.595±0.053
MAD-GAN	0.569±0.058	0.914±0.098	0.694±0.017	0.661±0.047	0.527±0.014	**0.987±0.022**	0.687±0.011	0.546±0.029	0.508±0.004	**0.992±0.008**	0.672±0.003	0.585±0.034
GANomaly	0.660±0.015	0.931±0.038	0.773±0.009	0.771±0.031	0.616±0.105	0.864±0.113	0.719±0.021	0.640±0.073	0.619±0.062	0.880±0.081	0.727±0.024	0.705±0.051
BeatGAN	0.659±0.033	0.904±0.064	0.762±0.015	0.772±0.029	0.570±0.029	0.932±0.060	0.707±0.012	0.649±0.037	0.589±0.043	0.938±0.082	0.724±0.014	0.688±0.018
ECGGAN	**0.775±0.018**	0.936±0.045	**0.848±0.015**	**0.829±0.016**	**0.851±0.048**	0.911±0.020	**0.880±0.026**	**0.872±0.032**	**0.792±0.035**	0.926±0.038	**0.854±0.014**	**0.835±0.019**

4.2 Effectiveness

In Tabel 2, we summarize the overall experimental results measured in terms of precision, recall, F1-score, AUC on two real-world datasets and the mixed-set. We can observe that ECGGAN outperforms all baselines with a higher precision, F1-score, and AUC. Based on these results, we have the following findings.

Comparison with conventional feature extraction-based methods. Our method significantly outperforms SVM and KNN in terms of precision, F1-score, and AUC, with at least 22.5%, 16.8% and 18.4% improvement respectively across all datasets. Both SVM and KNN methods classify almost all ECG as abnormal, that is why the precision is close to 0.5 and recall is close to 1. On the contrary, ECGGAN has the advantage of capturing complex normal pattern with a more accurate performance for ECG anomaly detection.

Comparison with general classification-based methods. We construct the CNN model with a network similar to the anomaly detection module in ECGGAN for fairness. We can observe that ECGGAN outperforms CNN model on all datasets. Specifically, precision, F1-score, and AUC increase by 8.2%, 5.4% and 6.5% respectively on AIWIN, which fully demonstrates the effectiveness of our method on ECG anomaly detection task. Different from CNN which directly extracts features for anomaly detection, our method can better detect ECG anomaly by learning the periodic and fine-grained normal pattern in the ECG.

Comparison with general reconstruction-based methods. Our method produces significantly better detection results than the general reconstruction-based methods. As shown in Tabel 2, we improve 17.4%, 12.1% and 13.7% on precision, F1-score, and AUC. Among all the general reconstruction-based methods (i.e. LSTM, VAE, GAN, MAD-GAN and GANomaly, BeatGAN), GANomaly performs the best with more parameters. Even comparing to GANomaly, we achieve 11.5%, 7.5% and 5.8% improvement on precision, F1-score and AUC, respectively. Since general reconstruction-based methods are not specifically designed for detecting ECG anomaly, they misidentify most normal fluctuations (beats in ECG) as anomalies, which results in a high recall and low precision. In addition, even with the special design for beat, BeatGAN only performs anomaly detection in each beat rather than the entire ECG, so could not detect anomalies that appear within several beats or across beats. In contrast, ECGGAN is designed with consideration of the periodicity and regularity of the ECG data, conducts a conditional GAN to capture the unique characteristics of each lead and defines a residual matrix to retain the global information at ECG-level.

4.3 Ablation Studies of ECGGAN

To further evaluate the effectiveness of each component of ECGGAN, we gradually exclude the components to observe how the model performance degrades. First, to verify the effectiveness of lead constraints in the reconstruction process, we remove the constraints and directly generate beats for the reconstruction of 12-lead ECG. Second, we mask the G-gen loss and G-feature loss, replacing them with zero to study the importance of the regularization terms. Finally, we replace anomaly detection module by calculating the error between the original ECG and the reconstruction directly to study the effectiveness of the anomaly detection module of ECGGAN. Table 3 shows the results of ablation experiments.

w/o Lead Constraint. We can observe that the F1-score decreases on all three datasets after removing lead constraints from the reconstruction module, even if the improvement of AUC is not obvious on the mixed-set due to different data sources. The results indicate that the lead constraints can better capture the normal pattern in ECG and improve performance by learning the unique characteristics of each lead.

w/o Gen & Feature Loss. After removing the G-gen loss and G-feature loss, the recall on CPSC and the precision on AIWIN both decrease by 2.5%, and the F1-score also decreases by about 1%. The results confirm that the addition of regularization terms can better retain the specific information of different patients, which establishes the explicit association between the original ECG and its reconstruction, as well as the implicit association in hidden space.

w/o Anomaly Detection Module. We can observe that the results are affected badly on all datasets after removing anomaly detection module, with F1-scores decreasing by 9.2% to 17.4% and AUC decreasing by 10.8% to 21.8%. This is because ECGGAN w/o anomaly detection module requires the accumulated errors to reach a relatively large number in order to be detected, which is hard to detect minor anomalies.

4.4 Anomaly Interpretation

We demonstrate the interpretability results of ECGGAN at beat-level and ECG-level by locating the sites and leads of anomalies.

Beat-level Anomaly Detection and Localization. We first show the visualization results at beat-level in Figure 5. Specifically, for normal beat, we can observe that the reconstructed beat (shown in red) is very similar to the original beat (shown in black), and thus it is identified as normal. For abnormal beat, the original beat deviates significantly from the reconstructed beat with normal pattern, especially at the red and yellow areas (shown in the circled region), which reflect the sites of anomalies.

Table 3: Performance of ECGGAN without specified component

Method	CPSC				AIWIN				Mixed-Set			
	Precision	Recall	F1-score	AUC	Precision	Recall	F1-score	AUC	Precision	Recall	F1-score	AUC
w/o LC	**0.779±0.046**	0.884±0.044	0.828±0.037	0.813±0.046	0.831±0.026	0.901±0.029	0.865±0.010	0.857±0.011	0.812±0.021	0.897±0.037	0.852±0.021	**0.842±0.021**
w/o GFL	0.775±0.016	0.910±0.038	0.837±0.016	0.818±0.016	0.825±0.082	**0.922±0.062**	0.871±0.033	0.856±0.039	**0.819±0.049**	0.888±0.041	0.852±0.019	0.841±0.027
w/o ADM	0.630±0.037	**0.944±0.050**	0.756±0.021	0.721±0.043	0.586±0.058	0.888±0.086	0.706±0.017	0.654±0.042	0.5854±0.034	**0.947±0.024**	0.724±0.022	0.669±0.066
ECGGAN	0.775±0.018	0.936±0.045	**0.848±0.015**	**0.829±0.016**	**0.851±0.048**	0.911±0.020	**0.880±0.026**	**0.872±0.032**	0.792±0.035	0.926±0.038	**0.854±0.014**	0.835±0.019

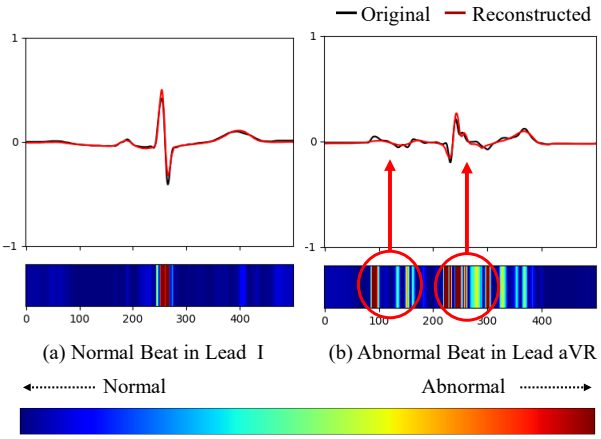
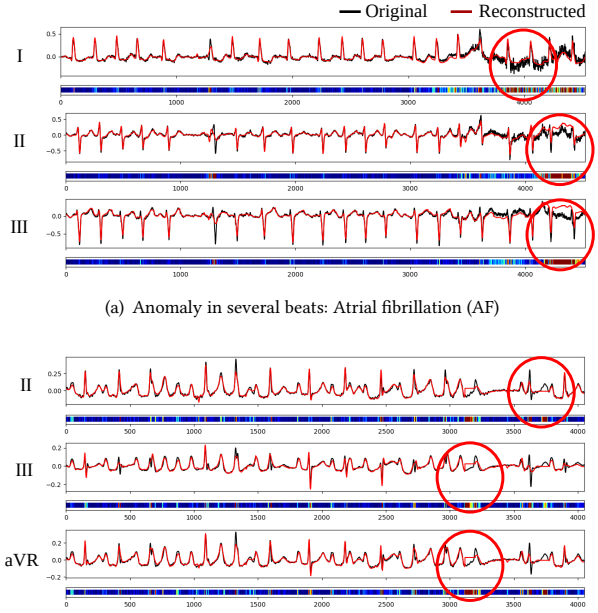


Figure 5: Anomaly detection and localization at beat-level

ECG-level Anomaly Detection and Localization. Figure 6 illustrates the anomaly detection and localization at ECG-level with two cases. The first case is Atrial Fibrillation (AF), which is a typical disease in the second category mentioned in Section 1 where the anomalies are present only within several beats, as shown in the upper of Figure 6(a), where the original ECG is shown in black, and the reconstructed ECG is shown in red. We can observe that the anterior beats of the original ECG are normal with a well fitting to the reconstructed ECG, while the posterior abnormal parts significantly deviate from the reconstructed ECG with normal pattern. And we automatically visualize the abnormal sites in specific lead as red by the anomaly interpretation module, shown in the red circles in the figure. In this way, ECGGAN can pinpoint the specific sites and leads of the abnormal part of the patient's ECG.

The second is Premature Atrial Contraction (PAC), which is the typical disease of the third category mentioned in Section 1, where the anomalies are reflected cross multiple beats, as shown in the bottom of Figure 6(b). First, we can notice that the intervals of beats are irregular, especially the intervals in the posterior part are far beyond normal. For the beats with normal morphology and interval, the original ECG fits well with the reconstructed ECG, as shown in the anterior part in the figure. But the posterior parts obviously deviate from the reconstructed ECG, identified as anomalies, and we automatically visualize these abnormal sites as red by the anomaly interpretation module, as shown in the red circles of the figure. The above two cases indicate that ECGGAN can accurately detect the anomalies at ECG-level and enables us to focus on the locations and leads of anomalies by visualizing the deviation between the original ECG and the reconstructed ECG.



(a) Anomaly in several beats: Atrial fibrillation (AF)

(b) Anomaly cross beats: Premature atrial contraction (PAC)

Figure 6: Anomaly detection and localization at ECG-level

5 CONCLUSIONS

In this paper, we study the characteristics of ECG data with regularity and periodicity and propose a novel ECG anomaly detection framework based on the reconstruction of periodic metadata, called ECGGAN. Taking leads of ECG as the constraints, ECGGAN captures the unique characteristics of each lead and learns the universal patterns from representative normal data. Based on our restoring mechanism that generates the complete reconstructed ECG signal, the anomaly detection module of ECGGAN can exploit lead information and the relationships between beats in the whole ECG to detect anomalies beyond the beat-level. An extensive experimental study on real datasets confirms that our ECGGAN achieves superior performance compared to existing methods in terms of precision, recall and F1-score. Moreover, ECGGAN can locate the specific sites and leads of the anomalies by visualizing the deviation between the original ECG and the reconstructed ECG, which helps to provide an interpretable result for clinical diagnosis.

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China (62072033) and the Fundamental Research Funds for the Central Universities. Zhaojing Luo's work is supported by Singapore Ministry of Education Academic Research Fund Tier 3 under MOE's official grant number MOE2017-T3-1-007.

REFERENCES

[1] AIWIN 2021. *AIWIN (Autumn) – ECG Intelligent Diagnosis Competition by SID Medical and Fudan University Zhongshan Hospital.* http://ailab.aiwin.org.cn/competitions/64#learn_the_details

[2] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. 2018. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*. Springer, 622–637.

[3] Mohamad Mahmoud Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and Ronald R Yager. 2016. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences* 345 (2016), 340–354.

[4] Erick A Perez Alday, Annie Gu, Amit J Shah, Chad Robichaux, An-Kwok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, et al. 2020. Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. *Physiological measurement* 41, 12 (2020), 124003.

[5] Subhi J Al'Aref, Khalil Anchouche, Gurpreet Singh, Piotr J Slomka, Kranthi K Kolli, Amit Kumar, Mohit Pandey, Gabriel Maliakal, Alexander R Van Rosendael, Ashley N Beecy, et al. 2019. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal* 40, 24 (2019), 1975–1986.

[6] Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*. PMLR, 214–223.

[7] Zachi I Attia, Suraj Kapa, Francisco Lopez-Jimenez, Paul M McKie, Dorothy J Ladewig, Gaurav Satam, Patricia A Pellikka, Maurice Enriquez-Sarano, Peter A Noseworthy, Thomas M Munger, et al. 2019. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature medicine* 25, 1 (2019), 70–74.

[8] Swati Banerjee and Madhuchhanda Mitra. 2013. Application of cross wavelet transform for ECG pattern analysis and classification. *IEEE transactions on instrumentation and measurement* 63, 2 (2013), 326–333.

[9] Igor Calzada. 2022. Citizens' data privacy in China: The state of the art of the Personal Information Protection Law (PIPL). *Smart Cities* 5, 3 (2022), 1129–1150.

[10] Shaojie Chen, Zhaopeng Meng, and Qing Zhao. 2018. Electrocardiogram recognization based on variational AutoEncoder. In *Machine Learning and Biometrics*. IntechOpen.

[11] Xuanhao Chen, Liwei Deng, Feiteng Huang, Chengwei Zhang, Zongquan Zhang, Yan Zhao, and Kai Zheng. 2021. Daemon: Unsupervised anomaly detection and interpretation for multivariate time series. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2225–2230.

[12] Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. 2020. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 492–502.

[13] Zhicheng Cui, Wenlin Chen, and Yixin Chen. 2016. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995* (2016).

[14] Aeatures Dallali, A Kachouri, and M Samet. 2011. Classification of cardiac arrhythmia using wt, hrv, and fuzzy c-means clustering. *Signal Processing: An Int. J.(SPJI)* 5, 3 (2011), 101–109.

[15] A Dallali, A Kachouri, and M Samet. 2011. Fuzzy c-means clustering, Neural Network, WT, and HRV for classification of cardiac arrhythmia. *ARPN Journal of Engineering and Applied Sciences* 6, 10 (2011), 2011.

[16] P De Chazal, BG Celler, and RB Reilly. 2000. Using wavelet coefficients for the classification of the electrocardiogram. In *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*, Vol. 1. IEEE, 64–67.

[17] Ailin Deng and Bryan Hooi. 2021. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4027–4035.

[18] Zahra Ebrahimi, Mohammad Loni, Masoud Daneshtalab, and Arash Gharehbaghi. 2020. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X* 7 (2020), 100033.

[19] Anubhav Garg and Michael H Lehmann. 2013. Prolonged QT interval diagnosis suppression by a widely used computerized ECG analysis system. *Circulation: Arrhythmia and Electrophysiology* 6, 1 (2013), 76–83.

[20] Thomas Gaziano, K Srinath Reddy, Fred Paccaud, Sue Horton, and Vivek Chaturvedi. 2006. Cardiovascular disease. *Disease Control Priorities in Developing Countries. 2nd edition* (2006).

[21] Alexander Geiger, Dongyu Liu, Sarah Alnegheimish, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. 2020. TadGAN: Time series anomaly detection using generative adversarial networks. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 33–43.

[22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[23] Patrick S Hamilton and Willis J Tompkins. 1986. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE transactions on biomedical engineering* 12 (1986), 1157–1165.

[24] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. 2019. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine* 25, 1 (2019), 65–69.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[26] Steven A Hicks, Jonas L Isaksen, Vajira Thambawita, Jonas Ghouse, Gustav Ahlberg, Allan Linneberg, Niels Grarup, Inga Strümke, Christina Ellervik, Morten Salling Olesen, et al. 2021. Explaining deep neural networks for knowledge discovery in electrocardiogram analysis. *Scientific reports* 11, 1 (2021), 10949.

[27] Shenda Hong, Yanbo Xu, Alind Khare, Satria Priambada, Kevin Maher, Alaa Aljiffry, Jimeng Sun, and Alexey Tumanov. 2020. Holmes: health online model ensemble serving for deep learning models in intensive care units. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1614–1624.

[28] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, and Jimeng Sun. 2020. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in biology and medicine* 122 (2020), 103801.

[29] Borui Hou, Jianyong Yang, Pu Wang, and Ruqiang Yan. 2019. LSTM-based auto-encoder model for ECG arrhythmias classification. *IEEE Transactions on Instrumentation and Measurement* 69, 4 (2019), 1232–1240.

[30] Hamid Khorrami and Majid Moavenian. 2010. A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Expert systems with Applications* 37, 8 (2010), 5751–5757.

[31] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[32] Richard Klabunde. 2011. *Cardiovascular physiology concepts*. Lippincott Williams & Wilkins.

[33] Joon-Myoung Kwon, Ki-Hyun Jeon, Hyue Mee Kim, Min Jeong Kim, Sung Min Lim, Kyung-Hee Kim, Pil Sang Song, Jinsik Park, Rak Kyeong Choi, and Byung-Hee Oh. 2020. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *EP Europace* 22, 3 (2020), 412–419.

[34] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. 1995. Detection of ECG characteristic points using wavelet transforms. *IEEE Transactions on biomedical Engineering* 42, 1 (1995), 21–28.

[35] Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. 2019. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In *International conference on artificial neural networks*. Springer, 703–716.

[36] Hongzu Li and Pierre Boulanger. 2020. A survey of heart anomaly detection using ambulatory Electrocardiogram (ECG). *Sensors* 20, 5 (2020), 1461.

[37] Taiyong Li and Min Zhou. 2016. ECG classification using wavelet packet entropy and random forests. *Entropy* 18, 8 (2016), 285.

[38] Zheng Jye Ling, Quoc Trung Tran, Ju Fan, Gerald CH Koh, Thi Nguyen, Chuen Seng Tan, James WL Yip, and Meihui Zhang. 2014. GEMINI: An integrative healthcare analytics system. *Proceedings of the VLDB Endowment* 7, 13 (2014), 1766–1771.

[39] Feifei Liu, Chengyu Liu, Lina Zhao, Xiangyu Zhang, Xiaoling Wu, Xiaoyan Xu, Yulin Liu, Caiyun Ma, Shoushui Wei, Zhiqiang He, et al. 2018. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics* 8, 7 (2018), 1368–1373.

[40] Pengfei Liu, Xiaoming Sun, Yang Han, Zhishuai He, Weifeng Zhang, and Chenxu Wu. 2022. Arrhythmia classification of LSTM autoencoder based on time series anomaly detection. *Biomedical Signal Processing and Control* 71 (2022), 103228.

[41] Yang Liu, Runnan He, Kuanquan Wang, Qince Li, Qiang Sun, Na Zhao, and Henggui Zhang. 2019. Automatic detection of ECG abnormalities by using an ensemble of deep residual networks with attention. In *Machine learning and medical engineering for cardiovascular health and intravascular imaging and computer assisted stenting*. Springer, 88–95.

[42] J Murray Longmore, Murray Longmore, Ian Wilkinson, Edward Davidson, Alexander Foulkes, and Ahmad Mafi. 2010. *Oxford handbook of clinical medicine*. Oxford University Press.

[43] Zhaojing Luo, Shaofeng Cai, Gang Chen, Jinyang Gao, Wang-Chien Lee, Kee Yuan Ngiam, and Meihui Zhang. 2019. Improving data analytics with fast and adaptive regularization. *IEEE Transactions on Knowledge and Data Engineering* 33, 2 (2019), 551–568.

[44] Zhaojing Luo, Shaofeng Cai, Can Cui, Beng Chin Ooi, and Yang Yang. 2021. Adaptive knowledge driven regularization for deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 8810–8818.

[45] Zhaojing Luo, Shaofeng Cai, Jinyang Gao, Meihui Zhang, Kee Yuan Ngiam, Gang Chen, and Wang-Chien Lee. 2018. Adaptive lightweight regularization tool for complex analytics. In *Proceedings of the 34th International Conference on Data Engineering*. 485–496.

[46] Zhaojing Luo, Shaofeng Cai, Yatong Wang, and Beng Chin Ooi. 2023. Regularized Pairwise Relationship Based Analytics for Structured Data. *Proc. ACM Manag. Data* 1, 1, Article 82 (2023), 27 pages.

[47] Zhaojing Luo, Sai Ho Yeung, Meihui Zhang, Kaiping Zheng, Lei Zhu, Gang Chen, Feiyi Fan, Qian Lin, Kee Yuan Ngiam, and Beng Chin Ooi. 2021. MLCask: Efficient management of component evolution in collaborative data analytics pipelines. In *Proceedings of the 37th International Conference on Data Engineering*. 1655–1666.

[48] Fenglong Ma, Muchao Ye, Junyu Luo, Cao Xiao, and Jimeng Sun. 2021. Advances in mining heterogeneous healthcare data. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 4050–4051.

[49] Roshan Joy Martis, U Rajendra Acharya, and Lim Choo Min. 2013. ECG beat classification using PCA, LDA, ICA and discrete wavelet transform. *Biomedical Signal Processing and Control* 8, 5 (2013), 437–448.

[50] Pankaj Mathur, Shweta Srivastava, Xiaowei Xu, and Jawahar L Mehta. 2020. Artificial intelligence, machine learning, and cardiovascular disease. *Clinical Medicine Insights: Cardiology* 14 (2020), 1179546820927404.

[51] W Mautgreve, Chr Klunker, and Chr Zywietz. 1989. HES EKG expert-an expert system for comprehensive ECG analysis and teaching. In *[1989] Proceedings. Computers in Cardiology*. IEEE, 77–80.

[52] FA Naima and AA Timemy. 2009. Neural network based classification of myocardial infarction: a comparative study of Wavelet and Fourier transforms. *Pattern Recognition* (2009), 337–351.

[53] Julien Oster, Jemma C Hopewell, Klemen Ziberna, Rohan Wijesurendra, Christian F Camm, Barbara Casadei, and Lionel Tarassenko. 2020. Identification of patients with atrial fibrillation: a big data exploratory analysis of the UK Biobank. *Physiological Measurement* 41, 2 (2020), 025001.

[54] Boris Pyakillya, Natasha Kazachenko, and Nikolay Mikhailovsky. 2017. Deep learning for ECG classification. In *Journal of physics: conference series*, Vol. 913. IOP Publishing, 012004.

[55] General Data Protection Regulation. 2018. General data protection regulation (GDPR). *Intersoft Consulting, Accessed in October* 24, 1 (2018).

[56] Jonathan Rubin, Saman Parvaneh, Asif Rahman, Bryan Conroy, and Saeed Babaeizadeh. 2017. Densely connected convolutional networks and signal quality analysis to detect atrial fibrillation using short single-lead ECG recordings. In *2017 Computing in cardiology (cinc)*. IEEE, 1–4.

[57] Ridhi Saini, Namita Bindal, and Puneet Bansal. 2015. Classification of heart diseases from ECG signals using wavelet transform and kNN classifier. In *International Conference on Computing, Communication & Automation*. IEEE, 1208–1215.

[58] Giovanna Sannino and Giuseppe De Pietro. 2018. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Generation Computer Systems* 86 (2018), 446–455.

[59] Stephen W Smith, Brooks Walsh, Ken Grauer, Kyuhyun Wang, Jeremy Rapin, Jia Li, William Fennell, and Pierre Taboulet. 2019. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *Journal of electrocardiology* 52 (2019), 88–95.

[60] M Llamedo Soria and JP Martinez. 2007. An ECG classification model based on multilead wavelet transform features. In *2007 Computers in Cardiology*. IEEE, 105–108.

[61] Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. 2020. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics* 25, 5 (2020), 1519–1528.

[62] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2828–2837.

[63] Nitish V Thakor and Y-S Zhu. 1991. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE transactions on biomedical engineering* 38, 8 (1991), 785–794.

[64] Paul van Gent, Haneen Farah, Nicole van Nes, and Bart van Arem. 2019. Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project. *Journal of Open Research Software* 7, 1 (2019).

[65] Paul Van Gent, Haneen Farah, Nicole Van Nes, and Bart Van Arem. 2019. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation research part F: traffic psychology and behaviour* 66 (2019), 368–378.

[66] Tao Wang, Changhua Lu, Yining Sun, Mei Yang, Chun Liu, and Chunsheng Ou. 2021. Automatic ECG classification using continuous wavelet transform and convolutional neural network. *Entropy* 23, 1 (2021), 119.

[67] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.

[68] Zhaohan Xiong, Martin K Stiles, and Jichao Zhao. 2017. Robust ECG signal classification for detection of atrial fibrillation using a novel neural network. In *2017 Computing in Cardiology (CinC)*. IEEE, 1–4.

[69] Kai Yang, Zhaojing Luo, Jinyang Gao, Junfeng Zhao, Beng Chin Ooi, and Bing Xie. 2021. LDA-Reg: Knowledge Driven Regularization Using External Corpora. *IEEE Transactions on Knowledge and Data Engineering* 34, 12 (2021), 5840–5853.

[70] Siti Agrippina Alodia Yusuf and Risanuri Hidayat. 2019. MFCC feature extraction and KNN classification in ECG signals. In *2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*. IEEE, 1–5.

[71] Chaohe Zhang, Xu Chu, Liantao Ma, Yinghao Zhu, Yasha Wang, Jiangtao Wang, and Junfeng Zhao. 2022. M3Care: Learning with Missing Modalities in Multimodal Healthcare Data. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2418–2428.

[72] Leigang Zhang, Hu Peng, and Chenglong Yu. 2010. An approach for ECG classification based on wavelet feature extraction and decision tree. In *2010 International Conference on Wireless Communications & Signal Processing (WCSP)*. IEEE, 1–4.

[73] Qibin Zhao and Liqing Zhang. 2005. ECG feature extraction and classification using wavelet transform and support vector machines. In *2005 International Conference on Neural Networks and Brain*, Vol. 2. IEEE, 1089–1092.

[74] Yunxiang Zhao, Jinyong Cheng, Ping Zhang, and Xueping Peng. 2020. ECG classification using deep CNN improved by wavelet transform. *Computers, Materials and Continua* (2020).

[75] Bin Zhou, Shenghua Liu, Bryan Hooi, Xueqi Cheng, and Jing Ye. 2019. BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series.. In *IJCAI*. 4433–4439.

[76] Sophia H Zhou, Eric D Helfenbein, James M Lindauer, Richard E Gregg, and Dirk Q Feild. 2009. Philips QT interval measurement algorithms for diagnostic, ambulatory, and patient monitoring ECG applications. *Annals of Noninvasive Electrocardiology* 14 (2009), S3–S8.